

Approximate explicit model predictive control via piecewise nonlinear system identification[☆]

Van-Vuong Trinh^{a,b,*}, Mazen Alamir^a, Patrick Bonnay^b

^aUniv. Grenoble Alpes, CNRS, GIPSA-lab, F-38000 Grenoble, France

^bUniv. Grenoble Alpes, CEA, INAC-SBT, F-38000 Grenoble, France

Abstract

This article presents an identification methodology to capture general relationships, with application to piecewise nonlinear approximations of model predictive control for constrained (non)linear systems. The mathematical formulation takes, at each iteration, the form of a constrained linear (or quadratic) optimization problem that is mathematically feasible as well as numerically tractable. The efficiency of the devised methodology is demonstrated via two industrial applications. Results suggest the possibility to achieve high approximate precision with limited number of regions, leading to a significant reduction in computation time when compared to the state-of-the-art implicit model predictive control solvers.

Keywords: Explicit model predictive control, Nonlinear approximation, System identification.

1. Introduction

Model predictive control (MPC) [1] is a model-based control method which involves the solution at each sampling period of a finite horizon optimal control problem subject to the system dynamics and constraints. This form of MPC, whose control actions are obtained by an online optimization, is referred to as implicit MPC. To reduce the online computational complexity of such an optimal control problem, two directions have been investigated. On the one hand, more efficient optimization strategies have been proposed in the aforementioned implicit MPC context. These approaches try to exploit as far as possible the problem structure and the possibility to use hot starts in order to reduce the number of iterations needed to reach a good solution (see e.g. [2] and the references therein). On the other hand, the so-called explicit MPC (EMPC) has been the subject of significant research efforts, where the control laws are available a priori, accordingly reducing an online optimization to a function evaluation. This contribution lies in this second direction with nonlinear systems as targeted domain of application.

For linear or hybrid systems, the explicit control laws can be exactly solved offline via parametric programming [3] [4]. As for nonlinear systems, deriving the true optimal nonlinear MPC (NMPC) control law is generally not possible and hence approximate approaches have to be followed. Following this direction, several techniques have been devised where the MPC

control laws are approximated using piecewise affine approximators [5], artificial neural networks [6] [7] and set memberships identification [8] [9]. Nevertheless, piecewise affine approaches normally lead to an excessive number of regions needed to represent the approximate control law while standard nonlinear approximators offer universal capabilities at the price of non-convex optimization schemes. Thus, approximating (or learning) the MPC control laws by piecewise nonlinear (PWNL) functions are recently of interest (see, for instance, [10]).

This paper suggests a practical computational methodology to EMPCs derived from learning data that incorporates implicit MPC solution and a problem-dependent regressor. In particular, each component of the control vector is identified independently as a PWNL map of the available regressor. A heuristic identification procedure to do so is proposed, which is an extension of the preliminary work [11]. Regarding complexity, the PWNL representations of the identified EMPCs allows to reduce the number of regions with respect to the piecewise affine approaches, especially for constrained NMPCs. Compared to the standard nonlinear structures, the methodology takes advantages of efficient computation of constrained linear (or quadratic) programming problems.

The preliminary version of this paper has been presented in [12]. The current work introduces additional details, including the improvement of algorithms, their numerical analysis as well as one more multivariable control application, namely, the constrained control of the compression station used in the cryogenic refrigerators.

The remainder of this paper is organized as follows. Section 2 recalls the identification of a nonlinear map proposed in [11]. The extension of this framework to the piecewise nonlinear maps is explained in Section 3 where an identification methodology is developed with application to EMPCs. Two

[☆]This work is partially supported by the French National Research Agency, task agreement ANR-13-SEED-0005.

*Corresponding author

Email addresses: van-vuong.trinh@gipsa-lab.fr (Van-Vuong Trinh), mazen.alamir@gipsa-lab.fr (Mazen Alamir), patrick.bonnay@cea.fr (Patrick Bonnay)

URL: <http://trinhvv.github.io> (Van-Vuong Trinh)

real control design problems are used in Section 5 to assess the ability of the proposed methodology to approximate the MPC control laws with rather small number of regions and the possibility to monitor the complexity as a function of the required precision. Moreover, the resulting computation time is compared to the one obtained by the state-of-the-art solver ACADO [13], showing the potential benefit from the explicit representation to extend the applicability of the MPC framework. Finally, Section 6 concludes the paper and gives hints for further investigations.

Notation

In the sequel, the following notation is extensively used: For $x \in \mathbb{R}^n$, its i -th component is x_i , the Euclidean norm is $\|x\|$, the weighted norm is $\|x\|_A = (x^T A x)^{1/2}$, the p -norm is $\|x\|_p$ with $p \geq 1$ be a real number, the absolute norm is $|x|$ when $n = 1$. For a discrete set, $\#\mathcal{X}$ means the cardinality, i.e. the number of elements of the set \mathcal{X} .

2. Design of A Nonlinear Approximator

In this section, we briefly recall the identification framework [11] for a class of nonlinear relationships. Several modifications with preliminary analysis are also introduced.

2.1. Problem Statement and Mathematical Formulation

Problem 1. Given the data $\mathcal{D} = \{(q(k), Z(k))\}_{k=1}^N$ where $Z \in \mathcal{Z} \subseteq \mathbb{R}^{n_z}$ is the regressor and $q \in [\underline{q}, \bar{q}] \subset \mathbb{R}$ is the output. Find a map $F : \mathbb{R}^{n_z} \rightarrow \mathbb{R}$ of the form:

$$F(Z) := \Gamma^{-1}(L^T Z) \quad (1)$$

where $L \in \mathbb{R}^{n_z}$, $\Gamma(\cdot)$ is strictly increasing and such that the following approximation holds:

$$q \approx F(Z) \quad (2)$$

The structure (1) is a Wiener-like model with a strictly increasing static output mapping whose inverse map can be parametrized using a finite function basis:

$$\Gamma(q) = \sum_{i=1}^{n_b} B^{(i)}(\xi(q)) \mu_i = B(\xi(q)) \mu; \quad \xi(q) = \frac{q - \underline{q}}{\bar{q} - \underline{q}} \quad (3)$$

where the basis functions are given by:

$$\{B^{(i)}\}_{i=1}^{n_b} := \{1\} \cup \{B_1^{(i)}\}_{i=2}^{n_m} \cup \{B_2^{(i)}\}_{i=1}^{n_m} \quad (4)$$

$$B_1^{(i)}(\eta) := (1 + \alpha_i) \frac{\eta}{1 + \alpha_i \eta}; \quad B_2^{(i)}(\eta) := \frac{\eta}{1 + \alpha_i(1 - \eta)} \quad (5)$$

The number of functions is $n_b = 2n_m$ while the coefficients α_i are given by $\alpha_i := e^{\beta(1-i)} - 1$ for some constant $\beta > 0$.

Denote $\mu \in \mathbb{R}^{n_b}$ and $L \in \mathbb{R}^{n_z}$ as the parameters of $F(Z)$. The total number of parameters is $n_p = n_b + n_z$. The basic idea of the above formulation is to solve Problem 1 by finding $\mu \in \mathbb{R}^{n_b}$ and $L \in \mathbb{R}^{n_z}$ such that the following approximation holds

$$B(\xi(q)) \mu \approx L^T Z \quad (6)$$

while guaranteeing the following inequalities

$$\left[\frac{dB}{d\eta}(\eta) \right] \mu \geq \epsilon \quad \forall \eta \in [0, 1] \quad (7)$$

given a positive constant ϵ .

The constraint (7) expresses the fact that $\Gamma(\cdot)$ has to be strictly increasing in order to guarantee the existence of the inverse map $\Gamma^{-1}(\cdot)$. This constraint can be transformed into a finite number of linear inequalities by defining a sufficiently dense grid of η over the interval $[0, 1]$ as

$$\eta = \{0 = \eta_1 < \eta_2 < \dots < \eta_{n_{\text{grid}}} = 1\} \quad (8)$$

and can be rewritten in the following compact form

$$\underbrace{\begin{bmatrix} \frac{dB^{(1)}}{d\eta}(\eta_1) & \dots & \frac{dB^{(n_b)}}{d\eta}(\eta_1) \\ \frac{dB^{(1)}}{d\eta}(\eta_2) & \dots & \frac{dB^{(n_b)}}{d\eta}(\eta_2) \\ \vdots & \ddots & \vdots \\ \frac{dB^{(1)}}{d\eta}(\eta_{n_{\text{grid}}}) & \dots & \frac{dB^{(n_b)}}{d\eta}(\eta_{n_{\text{grid}}}) \end{bmatrix}}_{[dB/d\eta(\eta)]} \cdot \underbrace{\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_{n_b} \end{bmatrix}}_{\mu} \geq \epsilon \cdot \underbrace{\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}}_1 \quad (9)$$

To summarize, the optimal parameters (μ, L) can be obtained by solving the following linear program (LP):

$$\min_{\mu, L} \max_{(q, Z) \in \mathcal{D}} \omega(q, Z) \cdot |B(\xi(q)) \mu - Z^T L| \quad (10a)$$

$$\text{s.t.} \quad \left[\frac{dB}{d\eta}(\eta) \right] \mu \geq \epsilon \cdot \mathbf{1} \quad (10b)$$

where the weight indicator $\omega(q, Z) : \mathbb{R} \times \mathbb{R}^{n_z} \rightarrow \mathbb{R}_{>0}$ is introduced in the cost function (10a) to enforce specific precision. This formulation obviously recalls the known *weighted norm approximation* where ∞ -norm is employed. Alternative formulation, based on the L_2 -norm, can also be adopted leading to a quadratic programming (QP) problem.

2.2. Preliminary Analysis

For convenience, the quadruple *identification parameters* $(n_m, \beta, \epsilon, \eta)$ is defined as follows:

- The parameter $n_m \geq 1$ is used to manipulate the degree of freedom;
- The parameter β is normally fixed as 0.5;
- The parameter ϵ is normally fixed as 1;
- The output grid η need to be sufficiently dense.

The strictly positive weight indicator $\omega(q, Z)$ can be small or large over some subsets of data while being equal to 1 for the remaining. The strict positivity of this weight is to avoid any drastically loose approximation as well as to guarantee the well-posedness of the formulation. In this paper, we utilize a simple form of the weight indicator $\omega(q, Z)$ as

$$\omega(q, Z) = \begin{cases} \rho_1 & \text{if } (q, Z) \in \mathcal{W}_1 \\ \rho_2 & \text{if } (q, Z) \in \mathcal{W}_2 \\ \vdots & \\ 1 & \text{otherwise} \end{cases} \quad (11)$$

where $\rho_i > 0$ are the constant weights corresponding to disjoint subspaces $\mathcal{W}_i \subset [q, \bar{q}] \times \mathcal{Z}$.

Once the identification parameters are set and the weight indicator is constructed, the *model parameters* (μ, L) can be obtained owing to the following feasibility assessment:

Proposition 1. *The LP (10) is feasible.*

Proof. According to (4) and (9), $[dB/d\eta(\eta)]$ is an $n_{\text{grid}} \times n_b$ matrix with the elements of the first column being zeros while the remaining being strictly positive. Hence, it is obvious that there exists μ , for instance, the one with $\mu_i > 0$ for all $i = 2, \dots, n_b$, such that the l.h.s. of (10b) is strictly positive. Thus, the fulfillment of the feasibility condition (10b) is guaranteed with any $\epsilon > 0$. \square

The identification residual of the identified model can be characterized as follows:

Proposition 2. *Given the map F whose model parameter, namely (μ, L) , is a feasible solution of the LP (10) with corresponding minimization cost*

$$J = \max_{(q, Z) \in \mathcal{D}} \omega(q, Z) \cdot |B(\xi(q))\mu - Z^T L| \quad (12)$$

The identification residual at any learning data point (q, Z) is such that:

$$\frac{|q - F(Z)|}{\bar{q} - \underline{q}} \leq \frac{1}{\omega(q, Z)} \cdot \frac{J}{\epsilon} \quad (13)$$

Proof. Consider any $(q, Z) \in \mathcal{D}$. Let us denote the approximate as $\hat{q} = F(Z)$, i.e. $B(\xi(\hat{q}))\mu = L^T Z$. Since the map $\Gamma(\cdot)$ is continuous differentiable, one clearly has

$$\begin{aligned} J &\geq \omega(q, Z) \cdot |B(\xi(q))\mu - L^T Z| \\ &= \omega(q, Z) \cdot |B(\xi(q))\mu - B(\xi(\hat{q}))\mu| \\ &\geq \min_{\xi \in [0, 1]} \left[\left[\frac{dB}{d\xi}(\xi) \right] \mu \right] \omega(q, Z) \cdot |\xi(\hat{q}) - \xi(q)| \\ &\geq \epsilon \cdot \omega(q, Z) \cdot \frac{|\hat{q} - q|}{\bar{q} - \underline{q}} \end{aligned}$$

which is equivalent to (13). \square

Proposition 2 implies that the desired fit would be obtained for a sufficiently small minimization cost. In other words, the approximations (2) and (6) are relevant.

3. Piecewise Nonlinear System Identification

The lack of universal property of the approximator (1) need to be underlined. In general, finding L and $\Gamma(\cdot)$ such that the approximation $\Gamma(q) \approx L^T Z$ holds might be impossible for many identification problems even for infinite number of parameters. This structural limitation can be overcome by using a number of submodels where each submodel valids over a region of the regression domain. An identification methodology, which is a generalization of piecewise affine frameworks, is proposed in this section to do so.

3.1. Problem Statement and Proposed Methodology

Problem 2. *Given the data $\mathcal{D} = \{(q(k), Z(k))\}_{k=1}^N$ where $Z \in \mathcal{Z} \subseteq \mathbb{R}^{n_z}$ is the regressor and $q \in [q, \bar{q}] \subset \mathbb{R}$ is the output. Find s maps $F_{(i)} : \mathbb{R}^{n_z} \rightarrow \mathbb{R}$ of the form:*

$$F_{(i)}(Z) := \Gamma_{(i)}^{-1}(L_{(i)}^T Z) \quad (14)$$

where $L_{(i)} \in \mathbb{R}^{n_z}$, the strictly increasing map $\Gamma_{(i)}(\cdot)$ is parametrized as

$$\Gamma_{(i)}(q) = B(\xi(q))\mu_{(i)}; \xi(q) = \frac{q - \underline{q}}{\bar{q} - \underline{q}} \quad (15)$$

and corresponding regions $\{\mathcal{R}_{(i)}\}_{i=1}^s$ forming a complete partition of \mathcal{Z} (i.e. $\bigcup_{i=1}^s \mathcal{R}_{(i)} = \mathcal{Z}$ and $\mathcal{R}_{(i)} \cap \mathcal{R}_{(j)} = \emptyset \forall i \neq j$) such that the following approximation holds:

$$q \approx \begin{cases} F_{(1)}(Z) & \text{if } Z \in \mathcal{R}_{(1)} \\ \vdots \\ F_{(s)}(Z) & \text{if } Z \in \mathcal{R}_{(s)} \end{cases} \quad (16)$$

where s is referred to as the model complexity.

Once the partition $\{\mathcal{R}_{(i)}\}_{i=1}^s$ is constructed, the corresponding classification $\{\mathcal{D}_{(i)}\}_{i=1}^s$ can be defined as

$$\mathcal{D}_{(i)} = \{(q, Z) \in \mathcal{D} : Z \in \mathcal{R}_{(i)}\} \quad (17)$$

According to Proposition 1, the submodel parameters $(\mu_{(i)}, L_{(i)})$ of $F_{(i)}(\cdot)$ are hereafter available by solving the following LP:

$$\min_{\mu_{(i)}, L_{(i)}} \max_{(q, Z) \in \mathcal{D}_{(i)}} \omega(q, Z) |B(\xi(q))\mu_{(i)} - Z^T L_{(i)}| \quad (18a)$$

$$\text{s.t.} \quad \left[\frac{dB}{d\eta}(\eta) \right] \mu_{(i)} \geq \epsilon \cdot \mathbf{1} \quad (18b)$$

The challenge is to find the minimal number s of regions $\mathcal{R}_{(i)}$ such that the following conditions hold:

1. The associated dataset is such that

$$\#\mathcal{D}_{(i)} \geq \kappa \quad \forall i = 1, \dots, s \quad (19)$$

where κ is the minimum required cardinality of each dataset. This lower bound is imposed to avoid overfit.

2. The associated identification residual is bounded by a given tolerance $\sigma > 0$, i.e.

$$\gamma_{(i)} = \max_{(q, Z) \in \mathcal{D}_{(i)}} \frac{|q - F_{(i)}(Z)|}{\bar{q} - \underline{q}} \leq \sigma \quad \forall i = 1, \dots, s \quad (20)$$

The structure (16) satisfying (20) will be referred to as an s -regions σ -error PWNL model. The following heuristic identification methodology is proposed to derive such a model, consisting of two stages:

1. *Domain partitioning:* At this stage, the partition is initially rough and is then refined at each step. Precisely, we seek partition $\{\mathcal{R}_{(i)}\}_{i=1}^r$ and associated classification $\{\mathcal{D}_{(i)}\}_{i=1}^r$, leading to a r -regions σ -error approximation with a *raw complexity* r for the desired fit. Generally, the lower σ , the finer resulting partition.

2. *Complexity reduction*: At this stage, the initial partition is the one resulted from the previous stage and is simplified at each step. Precisely, we aim to obtain an approximation with the *reduced complexity* $s < r$ by merging iteratively regions and datasets while maintaining the fitting level of σ -error.

The following sub-sections detail such stages.

3.2. Domain Partitioning

For convenience, all involved variables should be normalized. Hence, we restrict our attention to hyperrectangular regression domains $\mathcal{Z} = [0, 1] \times \dots \times [0, 1] \subset \mathbb{R}^{n_z}$ without loss of generality.

An iterative algorithm is proposed for the nonuniform hyperrectangular partitioning of PWNL function domains, i.e. each region $\mathcal{R}_{(i)}$ is a hyperrectangle, satisfying both (19) and (20). In addition, we utilize ν as the minimum allowed size of the hyperrectangles along each dimension. This lower bound is not only imposed by the machine precision but also can be tuned by the designer. The algorithm is summarized as follows:

1. *Initialization*: The algorithm starts with a rough initial partition which can be the whole domain, i.e. $r^{[0]} = 1$, $\mathcal{R}_{(1)}^{[0]} = \mathcal{Z}$ and $\mathcal{D}_{(1)}^{[0]} = \mathcal{D}$.
2. *Algorithm iterations*: At each iteration k , a hyperrectangle $\mathcal{R}_{(i)}^{[k]}$ with $\gamma_{(i)}^{[k]} > \sigma$ is split into two subhyperrectangles by an axis-orthogonal hyperplane that goes through the centroid of the dataset $\mathcal{D}_{(i)}^{[k]}$. The basic idea is to select the hyperplane leading to a maximal reduction in the identification residual while guaranteeing the size of the resulted subhyperrectangles and the cardinality of the associated datasets are not lower than ν and κ , respectively. In the sequel, the selected hyperplane will be replaced by a maximum-margin hyperplane that is parallel with it. Then, a new iteration of the algorithm is performed by increasing the complexity by one unit $r^{[k+1]} = r^{[k]} + 1$. Otherwise, the split fails when resulted subhyperrectangles and sub-datasets are too small for all split options.
3. *Termination*: The algorithm ends when the identification residual is not larger than σ or in the case of splitting failure.

The algorithm successfully terminates when, at the end of the iterations, a σ -error approximation is achieved while the size of all hyperrectangles and the cardinality of all datasets are not lower than ν and κ , respectively.

3.3. Complexity Reduction

Let us assume that the domain partitioning stage returns a partition $\{\mathcal{R}_{(i)}\}_{i=1}^r$, a classification $\{\mathcal{D}_{(i)}\}_{i=1}^r$ and the raw complexity r . If the model complexity r needed for data fit is overestimated, this number can be reduced by forcing the parameters of some partitions to be identical while maintaining the fitting level in a postprocessing step as follows:

1. *Initialization*: The algorithm starts with an admissible initial partition $\{\mathcal{R}_{(i)}\}_{i=1}^r$, i.e. $s^{[0]} = r$.

2. *Algorithm iterations*: At each iteration k , a nonempty subset $\mathcal{I} \subseteq \{1, \dots, s^{[k]}\}$ is computed satisfying the following conditions:

$$\begin{bmatrix} \mu_{(i)} \\ L_{(i)} \end{bmatrix} = \begin{bmatrix} \mu_{(j)} \\ L_{(j)} \end{bmatrix} \quad \forall i \neq j; i, j \in \mathcal{I} \quad (21a)$$

$$\gamma_{(i)} \leq \sigma \quad \forall i \in \mathcal{I} \quad (21b)$$

Then, a new iteration is performed by reducing the model complexity $s^{[k+1]} = s^{[k]} - \#\mathcal{I} + 1$. Moreover, one merges all the regions and datasets whose indexes belonging to \mathcal{I} .

3. *Termination*: The algorithm always successfully terminates and stops when the maximum number of iterations N_{iter} is reached.

The search for the subset \mathcal{I} at each iteration k will be detailed. Since searching for the subset \mathcal{I} over the whole set of regions $\{1, \dots, s^{[k]}\}$ is computationally expensive, the search over several regions with indexes belonging to a set $\mathcal{L} \subseteq \{1, \dots, s^{[k]}\}$ might be more practical. Once the set \mathcal{L} is chosen, its subset \mathcal{I} satisfying (21a) and (21b) is referred to as a *compatible subset*. The searching for the compatible subset \mathcal{I} is summarized as follows:

1. *Initialization*: The procedure start with an initial guess $\mathcal{I}^{[0]} = \mathcal{L}$.
2. *Algorithm iterations*: At each sub-iteration l , consider the following LP whose decision variables are $\{\mu_{(i)}, L_{(i)}\}_{i \in \mathcal{L}}$ and $\tilde{\mu}, \tilde{L}$:

$$\min \sum_{i \in \mathcal{L}} \max_{(q, Z) \in \mathcal{D}_{(i)}} \omega(q, Z) |B(\xi(q))\mu_{(i)} - Z^T L_{(i)}| \quad (22a)$$

$$\text{s.t.} \quad \left[\frac{dB}{d\eta}(\eta) \right] \mu_{(i)} \geq \epsilon \cdot \mathbf{1} \quad \forall i \in \mathcal{L} \quad (22b)$$

$$\text{and} \quad \left\| \begin{bmatrix} \mu_{(i)} \\ L_{(i)} \end{bmatrix} - \begin{bmatrix} \tilde{\mu} \\ \tilde{L} \end{bmatrix} \right\|_{\infty} = 0 \quad \forall i \in \mathcal{I}^{[l]} \quad (22c)$$

The feasibility of the LP (22) is straightforward from Proposition 1. The additional constraints (22c) are referred to as *compatibility constraints*, including $\#\mathcal{I}^{[l]}$ linear equalities. If the LP (22) results into the submodels with the following identification residual

$$\gamma_{(\mathcal{I}^{[l]})} = \max_{i \in \mathcal{I}^{[l]}} \gamma_{(i)} \quad (23)$$

which is larger than σ , it is an indication that the minimization cost of (22) is insufficiently small. This cost can be reduced by removing some of the compatibility constraints (22c). Since the sensitivities of (22) with respect to such constraints are represented by the corresponding Lagrange multipliers, namely $\{\lambda_{(i)}\}_{i \in \mathcal{I}}$, a common strategy is to iteratively remove the constraint associated with the largest Lagrange multiplier. This is equivalent to removing an element of \mathcal{I} , i.e.

$$\mathcal{I}^{[l+1]} = \mathcal{I}^{[l]} \setminus \{\arg\max_{i \in \mathcal{I}^{[l]}} \lambda_{(i)}\} \quad (24)$$

3. *Termination*: The procedure stops when $\gamma_{(\mathcal{I}^{[l]})} \leq \sigma$.

4. Approximate Explicit Model Predictive Control

4.1. Extension to Multi-Input Multi-Output MPC

Although the presented methodology is restricted to the multi-input single-output relationships, the extension to approximate an MPC control law is trivial by approximating each component of the control vector (see Sub-section 5.2 for example).

4.2. On the Stabilization

A crucial aspect of the resulting EMPC in the regulation problem is setpoints stabilization. On the one hand, the key properties to guarantee stabilization of such an approximate control law have been extensively studied in the literature. Example is [5] which indicates the boundedness of the state trajectory and its convergence to a neighborhood of the setpoint can be achieved by a sufficient small approximation error. A quantitative analysis on such an upper bound is presented in [8], however, is difficult to applied here due to the limitations of the identification approach. On the other hand, the efforts to establish asymptotic stability in a setpoint's neighborhood generally deploy the dual-mode design where a local stabilizing controller is used. In this paper, we impose a sufficiently high weight for the data point in a neighborhood of the setpoint to acquire an extremely high local accuracy. This parameter choice is based on the approximate stability result of discontinuous control laws (Theorem 1, Assumption 3, [14]).

4.3. Remarks on Weight Tuning

This can be generalized to the reference tracking problem. First, we extract from the learning data a discrete set of samples corresponding to stationary state of the closed loop. Second, in the identification procedure of control actuators, a typically high weight, namely ρ_{st} , is associated to such stationary data to obtain nearly zero steady errors of the closed-loop controlled by the EMPCs. Lastly, a moderate weight, namely ρ_{stab} , is associated to the data points in the neighborhood of the considered stationary data.

Note finally that a piecewise control law may results into undesired oscillations in the closed-loop response. This can be partially overcome by imposing a moderate weight, namely ρ_{sw} , for the data in the neighborhoods of the switching surfaces between computed regions.

5. Illustrative Examples

5.1. Application 1: Reference Tracking Control of A Stirling Engine Based Power Generation System

Consider the following system which belongs to a standalone hybrid Stirling engine / supercapacitor power plant (see Fig. 1) with $x_1 = \Omega$ being the shaft rotational speed, $x_2 = I_{red}$ and $x_3 = V_{red}$ being the rectified current and the rectified voltage while $x_4 = I_{Lfb}$ and $u = \alpha_{fb}$ being the output current and the

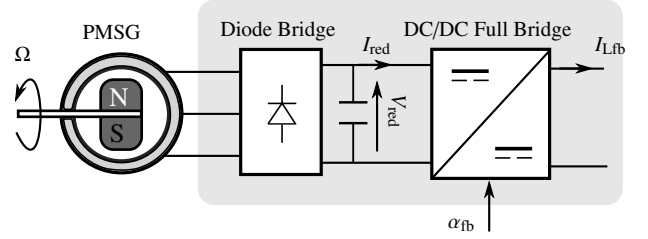


Figure 1: Schematic diagram of a power electronics module in thermodynamic power plants. The component named PMSG is permanent magnetic synchronous generator whose shaft is connected with a Stirling engine.

duty ratio of the DC/DC full bridge converter:

$$\begin{aligned} \dot{x}_1 &= -a_1 x_1 - a_3 x_2 + a_2 \\ \dot{x}_2 &= -a_4 x_2 + a_6 x_1 - a_7 x_3 \\ \dot{x}_3 &= a_8 (x_2 - k x_4 u) \\ \dot{x}_4 &= a_9 (-x_3^{st} + k x_3 u) \end{aligned} \quad (25)$$

The parameters of this model are $a_1 = 0.183$, $a_2 = 558.11$, $a_3 = 118.4453$, $a_4 = 9615.4$, $a_6 = 5101.1$, $a_7 = 641.02$, $a_8 = 425.53$, $a_9 = 6666.7$, $k = 0.5$ and $x_3^{st} = 50$. The system states should satisfy positivity constraints $x \geq 0$ since the used converter is not reversible. The control variable u is the duty ratio of this converter and has a strong saturation $u \in [0, 1]$. The sampling time is fixed as $T_s = 100\mu s$.

The control objective is to force the state x_4 to track a reference signal x_4^r while respecting the constraints. In [15], it has been shown that this is a challenging problem as (25) contains highly oscillatory modes that induce constraints violation if the latter are not explicitly addressed. These make the NMPC methodology particularly suitable. This can be achieved by minimizing the following cost:

$$\sum_{i=1}^{N_p} (\|x(k+i) - x^r(k)\|_Q^2 + \|u(k+i-1) - u^r(k)\|_R^2) + \rho \delta^2 \quad (26)$$

with the prediction horizon $N_p = 3$, the cost matrices $Q = I$, $R = 1$ and the weight penalizing constraint violations $\rho = 10^4$ while fulfilling the hard constraint $0 \leq u(k) \leq 1$, the soft constraints $[0 - \delta, 4.5 - 0.1\delta, 55 - \delta, 1 - \delta]^T \leq x(k) \leq [40 + \delta, 5.5 + 0.1\delta, 200 + \delta, 25 + \delta]^T$ where $\delta \geq 0$. The prediction simulator is under-sampled with a sampling period of $\tau_u = 1ms$.

The fast dynamics of the considered system would require the design of an EMPC controller. In order to do this, a data record is generated by simulating the closed-loop system with the implicit NMPC during a 10s scenario. This simulation scenario is based on the considered range of the reference as $x_4^r \in [7, 13]$ [A]. Hereafter, a learning data is extracted with the cardinality of $N = 24205$ including 48 stationary data points that are then used to build the instances

$$Z(k) = [x^T(k), (x^r(k))^T, (u^r(k))^T, (e(k))^T]^T; q(k) = u(k) \quad (27)$$

The reference tracking error $e(k) = x(k) - x^r(k)$ is incorporated into the regressor $Z(k)$ to further reduce the partition complexity

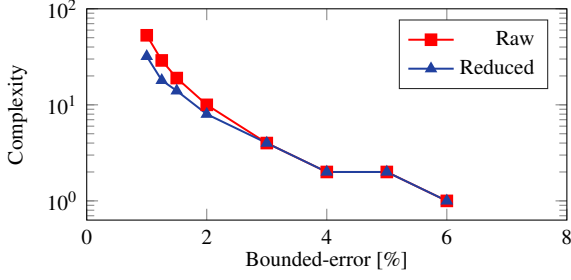


Figure 2: Trade-off curve.

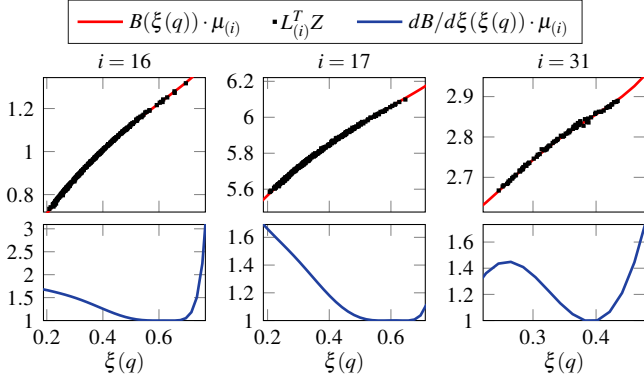


Figure 3: Nonlinearity and strict monotonicity of the i -th submodels corresponding to the 32-regions 1%-error EMPC. The black dots are the function evaluation for validation data while the red and the blue curves are the identified nonlinear function and its derivative, respectively.

of the explicit controller. The values $\underline{q} = 0.35$, $\bar{q} = 0.7$ are the range of u in the learning data.

Several approximate controllers have been identified in Matlab using the IBM's CPLEX with the same identification and partitioning parameters. The identification parameters have been chosen with $\beta = 0.5$, $\epsilon = 1$, $n_m = 10$ and a uniform grid η of $n_{\text{grid}} = 50$ points. The weight indicator is defined by imposing a high weight of $\rho_{\text{st}} = 10^2$ for stationary data, a moderate weight of $\rho_{\text{stab}} = 10$ for data in the neighborhood of stationary data, a lower weight of $\rho_{\text{sw}} = 2$ for data near the boundary of each hypercubes and the weight of 1 for the remaining. The partitioning parameters have been chosen as $\nu = 5 \times 10^{-2}$ and $\kappa = 50$.

Fig. 2 portrays a complexity assessment of identified controllers with different values of the bounded-error σ while the nonlinear characteristic of several i -th submodels corresponding to a 32-regions 1%-error EMPC are illustrated in Fig. 3. In Fig. 3, the red curves represent the identified maps $\Gamma_{(i)}^{-1}(\xi)$ which are strictly increasing as their derivatives (blue curves) are higher than $\epsilon = 1$. Although these maps are defined over the interval $\xi \in [0, 1]$, we only show the range with the presence of data points. The black dots illustrates the nonlinear nature of data.

The performance of the resulted EMPCs is evaluated by simulations and is compared with the closed-loop controlled by the

implicit NMPC. Fig. 4 shows the closed-loop response with different controllers: the implicit NMPC (green), the 32-regions 1%-error EMPC (red) and the 4-regions 3%-error EMPC (blue). Notably, there are switching between regions, especially in the transient duration. It is also clear that the tracking errors of those EMPCs are negligible.

Regarding real-time aspect, Fig. 5 illustrates the computation efficiency of the EMPCs. Their computation time vary between $8\mu\text{s}$ and $18\mu\text{s}$ which is slightly faster than that of the implicit NMPC delivered by ACADO and obviously suitable for this application.

5.2. Application 2: Control of A Compression Station

To illustrate the efficiency of the proposed framework for a multi-input multi-output case, we consider the constrained control for a warm compression station [18] (see Fig. 6). Variables to be regulated are the low pressure P^C and the high pressure P^H . The unmeasured flows M^H and M^C , respectively denote the inflow and outflow, are considered as the process disturbances. M_{C1} and M_{C2} respectively denote the flowrates handled by the first and second compressor which are variable frequency driven, M_{V1} represents the flowrate passing through the bypass valve V1 while M_{V2} and M_{V3} depict the flows generated by the charge valve V2 and discharge valve V3. These five variables are the manipulated variables of the process. Valves opening is obviously in the range $[0, 100]$ [%], the compressor C1 is limited to a speed in the range $[30, 53]$ [Hz] while the compressor C2 is manually set at a speed of 50 [Hz]. A simulator built up from the SimCryogenics Toolbox [19] is available.

In [18], the three variables $M_{\text{prod}} = M_{C1} + M_{C2} + M_{V1}$, $M_{\text{ch}} = M_{V2}$ and $M_{\text{dis}} = M_{V3}$ are introduced as the virtual actuators. Then, the virtual actuators value (in [kg/s]) will have to be converted into actual control actions for the system actuators (in [Hz] or [%]) by a conversion function. Then, a discrete-time linear model in the following form

$$x(k+1) = A \cdot x(k) + B \cdot u(k) + F \cdot d(k) \quad (28)$$

with the sampling time $T_s = 0.25\text{s}$ is synthesized for the control design. The system states, the control inputs and the disturbances are defined as the deviations around an operating point of interest, i.e.

$$x = \begin{bmatrix} \Delta P^H \\ \Delta P^C \end{bmatrix}; u = \begin{bmatrix} \Delta M_{\text{prod}} \\ \Delta M_{\text{ch}} \\ \Delta M_{\text{dis}} \end{bmatrix}; d = \begin{bmatrix} \Delta M^H \\ \Delta M^C \end{bmatrix} \quad (29)$$

The control objective is to regulate system states to the set-point $P^H = 16$ [bar], $P^C = 1.05$ [bar] under variable disturbances as well as input saturation. In addition, the low pressure is limited by a soft constraint $P^C \in [1, 1.1]$ [bar]. In [18], a MPC design is derived with the following cost

$$\sum_{i=1}^{N_p} (\|x(k+i)\|_Q^2 + \|u(k+i-1) - u^r(k)\|_R^2) \quad (30)$$

in which the control reference u^r is computed based on the estimated disturbances. A typically long prediction horizon of $N_p = 200$ has been used.

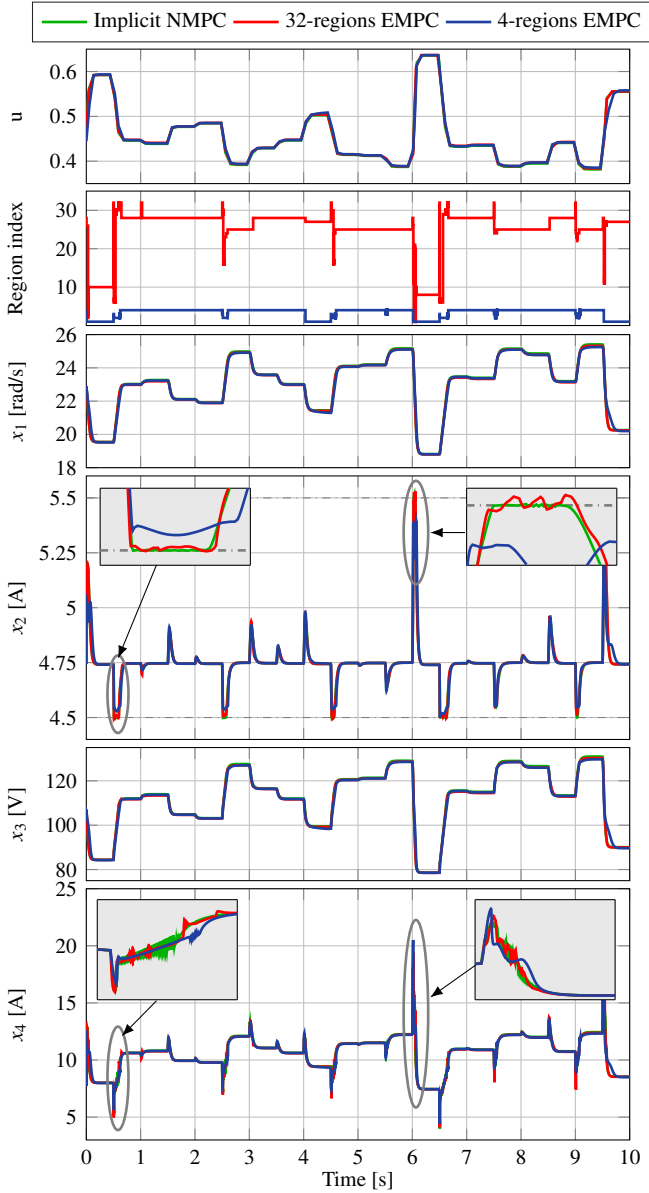


Figure 4: Closed-loop performance comparison. The constraint satisfaction are portrayed in the zooms on the evolution of x_2 .

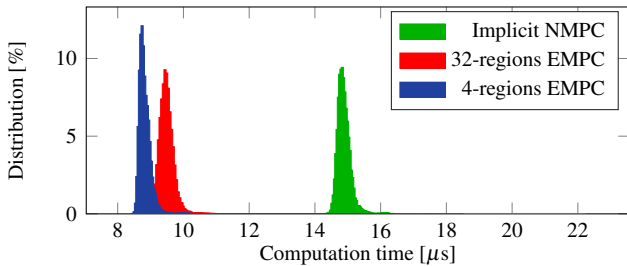


Figure 5: Computation time. The implicit NMPC was built using the ACADO code generation [16] with the qpOASES solver [17] while the EMPCs were implemented in C language. Platform: 2.6 GHz Intel(R) Core(TM) i7 and 16GB of RAM.

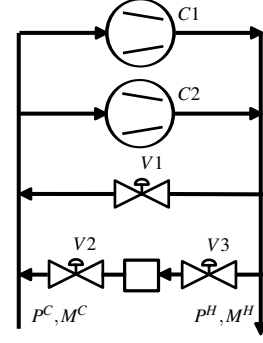


Figure 6: Schematic diagram of a warm compression station. Arrows represents the fluid's direction. The block between V2 and V3 valves represents the capacity buffer.

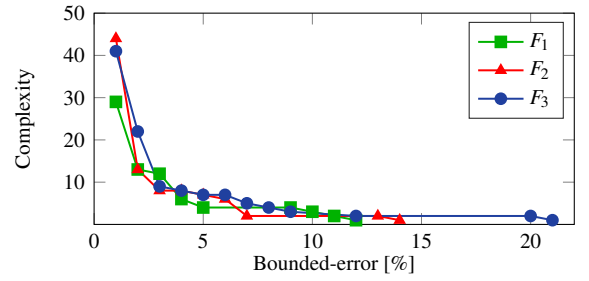


Figure 7: Trade-off curve.

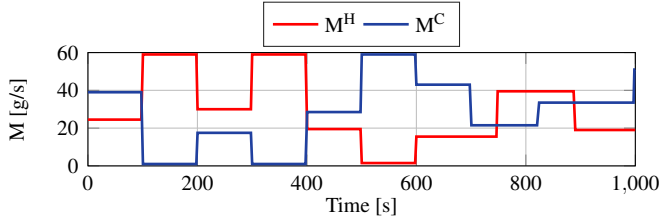
Our objective is to identify three maps F_i such that $q_i \approx F_i(Z)$, $i = 1, 2, 3$ where:

$$Z(k) = \begin{bmatrix} x(k) \\ u^r(k) \end{bmatrix}; q_i(k) = u_i(k) \quad (31)$$

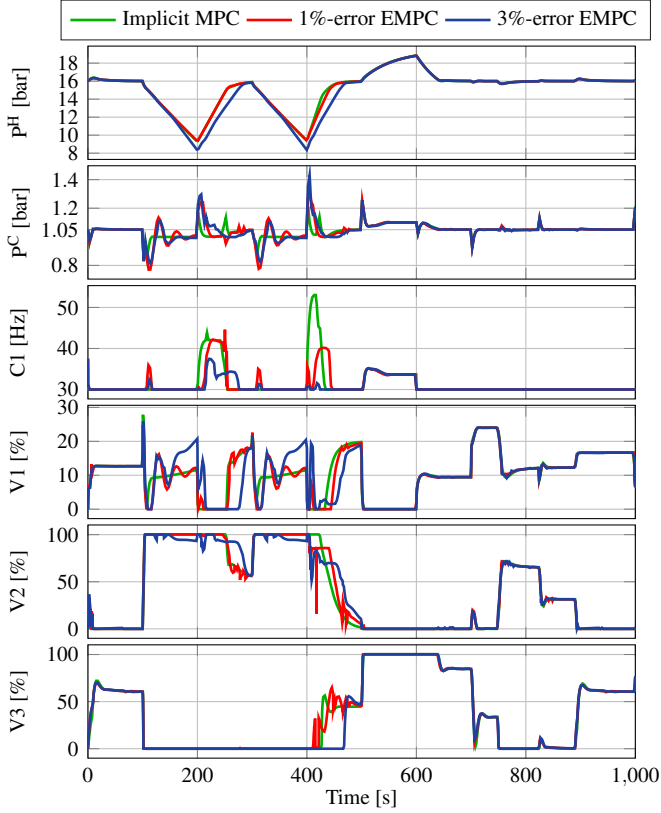
A data record over an interval of 25×10^3 s is available. It is hereafter extracted into a learning data with the cardinality of $N = 11633$ including 388 stationary data points. The identification parameters have been chosen with $\beta = 0.5$, $\epsilon = 1$, $n_m = 10$ and a nonuniform grid η of $n_{\text{grid}} = 100$ points. The weight indicator is defined by $\rho_{\text{st}} = 10^3$, $\rho_{\text{stab}} = 20$ and $\rho_{\text{sw}} = 5$. The partitioning parameters have been chosen as $\nu = 1 \times 10^{-2}$ and $\kappa = 10$.

Fig. 7 shows the trade off between the (reduced) complexity of the identified maps given different level of accuracy.

Fig. 8 depicts the closed-loop behaviours of the system in time under different controllers. The first 600s duration shows how the system reacts when disturbed by some excessive difference between the inflow and the outflow. It is worth noting that pressure stability is ensure and large disturbances are effectively handled even in this critical situation. With less difference between inflow and outflow in the last 400s duration, the performance of the implicit MPC and the identified EMPCs are obviously comparable. Notably, the pressures are returning to the setpoints without oscillations during such a normal situation.



(a) Inflow and outflow



(b) Controlled pressures and manipulated actuators

Figure 8: Closed-loop performance comparison. The depicted response of the pressures has an amount of performance degradation.

6. Conclusion and Future Work

In this paper, a new approach is proposed for explicit piecewise nonlinear presentation of approximate MPC control laws. Future research efforts will be devoted to different MPC settings such as output-feedback MPC, hybrid MPC or distributed MPC. Further developments include applying the methodology outside the MPC context, e.g. off-line computation of the moving horizon observers.

References

- [1] J. B. Rawlings and D. Q. Mayne, *Model predictive control: Theory and design*. Nob Hill Pub., 2009.
- [2] M. Diehl, H. J. Ferreau, and N. Haverbeke, "Efficient numerical methods for nonlinear mpc and moving horizon estimation," in *Nonlinear model predictive control*, pp. 391–417, Springer, 2009.

- [3] A. Bemporad, M. Morari, V. Dua, and E. N. Pistikopoulos, "The explicit linear quadratic regulator for constrained systems," *Automatica*, vol. 38, no. 1, pp. 3–20, 2002.
- [4] R. Oberdieck and E. N. Pistikopoulos, "Explicit hybrid model-predictive control: The exact solution," *Automatica*, vol. 58, pp. 152–159, 2015.
- [5] A. Grancharova and T. A. Johansen, *Explicit nonlinear model predictive control: theory and applications*, vol. 429. Springer Science & Business Media, 2012.
- [6] B. M. Åkesson and H. T. Toivonen, "A neural network model predictive controller," *Journal of Process Control*, vol. 16, no. 9, pp. 937–946, 2006.
- [7] G. Pin, M. Filippo, F. Pellegrino, G. Fenu, and T. Parisini, "Approximate model predictive control laws for constrained nonlinear discrete-time systems: analysis and offline design," *International Journal of Control*, vol. 86, no. 5, pp. 804–820, 2013.
- [8] M. Canale, L. Fagiano, and M. Milanese, "Efficient model predictive control for nonlinear systems via function approximation techniques," *IEEE Transactions on Automatic Control*, vol. 55, no. 8, pp. 1911–1916, 2010.
- [9] L. Fagiano, M. Canale, and M. Milanese, "Set membership approximation of discontinuous nonlinear model predictive control laws," *Automatica*, vol. 48, no. 1, pp. 191–197, 2012.
- [10] A. Domahidi, M. N. Zeilinger, M. Morari, and C. N. Jones, "Learning a feasible and stabilizing explicit model predictive control law by robust optimization," in *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, pp. 513–519, IEEE, 2011.
- [11] M. Alamir, "A new identification framework for off-line computation of moving-horizon observers," *IEEE Transactions on Automatic Control*, vol. 58, no. 7, pp. 1877–1882, 2013.
- [12] V.-V. Trinh, M. Alamir, P. Bonnay, and F. Bonne, "Explicit model predictive control via nonlinear piecewise approximations," in *10th IFAC Symposium on Nonlinear Control Systems*, 2016.
- [13] B. Houska, H. J. Ferreau, and M. Diehl, "Acado toolkit – an open-source framework for automatic control and dynamic optimization," *Optimal Control Applications and Methods*, vol. 32, no. 3, pp. 298–312, 2011.
- [14] P. O. Scokaert, D. Q. Mayne, and J. B. Rawlings, "Suboptimal model predictive control (feasibility implies stability)," *IEEE Transactions on Automatic Control*, vol. 44, no. 3, pp. 648–654, 1999.
- [15] M. A. Rahmani, M. Alamir, and D. Gualino, "Control strategy for an off-grid hybrid stirling engine/supercapacitor power generation system," in *American Control Conference (ACC), 2013*, pp. 4277–4282, IEEE, 2013.
- [16] R. Quirynen, M. Vukov, M. Zanon, and M. Diehl, "Autogenerating microsecond solvers for nonlinear mpc: A tutorial using acado integrators," *Optimal Control Applications and Methods*, vol. 36, no. 5, pp. 685–704, 2015.
- [17] H. J. Ferreau, C. Kirches, A. Potschka, H. G. Bock, and M. Diehl, "qpOASES: A parametric active-set algorithm for quadratic programming," *Mathematical Programming Computation*, vol. 6, no. 4, pp. 327–363, 2014.
- [18] F. Bonne, M. Alamir, and P. Bonnay, "Control of warm compression stations using mpc: Simulation and experimental results," in *ICEC/ICMC*, 2016.
- [19] F. Bonne, M. Alamir, C. Hoa, P. Bonnay, M. Bon-Mardion, and L. Monteiro, "A simulink library of cryogenic components to automatically generate control schemes for large cryorefrigerators," in *IOP Conference Series: Materials Science and Engineering*, vol. 101, p. 012171, IOP Publishing, 2015.